

0-781430

На правах рукописи



Зуев Денис Сергеевич

**МОДЕЛЬ И ПРОТОТИП ПРОГРАММНОЙ СИСТЕМЫ УПРАВЛЕНИЯ  
НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ ЭЛЕКТРОННОЙ БИБЛИОТЕКОЙ**

**Специальность 05.13.18 – Математическое моделирование,  
численные методы и комплексы программ**

**АВТОРЕФЕРАТ**  
**диссертации на соискание ученой степени**  
**кандидата технических наук**

Казань – 2010

Работа выполнена в Научно-исследовательском институте  
математики и механики им. Н. Г. Чеботарева  
Казанского государственного университета

**Научный руководитель:** доктор физико-математических наук,  
профессор,  
заслуженный деятель науки РТ  
Елизаров Александр Михайлович

**Официальные оппоненты:** доктор технических наук, профессор  
Захаров Вячеслав Михайлович

доктор физико-математических наук,  
профессор  
Соловьев Валерий Дмитриевич

**Ведущая организация:** Санкт-Петербургский государственный  
политехнический университет

Защита состоится «26» марта 2010 года в 14.30 часов на заседании  
Диссертационного совета Д 212.079.01 в Казанском государственном  
техническом университете им. А. Н. Туполева по адресу: 420111, Казань, ул.  
К. Маркса, д. 10.

С диссертацией можно ознакомиться в научной библиотеке Казанского  
государственного технического университета им. А. Н. Туполева. С  
авторефератом диссертации можно ознакомиться на сайте КГТУ  
им. А.Н. Туполева [www.kai.ru](http://www.kai.ru).

Автореферат разослан «25» февраля 2010 г.

Ученый секретарь диссертационного совета  
доктор физико-математических наук,

профессор



НАУЧНАЯ БИБЛИОТЕКА КГУ



0000590929

П. Г. Данилаев

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** Развитие вычислительной техники и сетевых технологий, появление новых носителей информации приводят к проникновению информационных технологий во все сферы человеческой жизнедеятельности, появляются новые классы информационных систем, новые виды сервиса. Необходимость современного информационного обеспечения научных исследований и учебного процесса, включения отечественной науки в мировое информационное пространство создали предпосылки для коренного изменения традиционных подходов к процессам информационного обеспечения науки и развития новых сервисов. Все это послужило толчком к созданию нового класса информационных систем, коими являются электронные библиотеки (ЭБ).

Сегодня в разработках коллекций информационных ресурсов электронных библиотек, обеспечении их поддержки и доступа к ним востребован весь спектр ключевых технологий управления информацией, которые используются в современных информационных системах. К тому же одной из основ развития информационных систем этого класса стали интернет и сопутствующие технологии. Поэтому разрабатываемые в настоящее время электронные библиотеки базируются на самых передовых достижениях веб-технологий (платформа XML, Symantec Web, языки описания онтологий, RDF/RDFS, OWL и др.), технологий баз данных, текстового поиска (полнотекстовый поиск, модели семантического поиска, новые подходы, ориентированные на текстовый поиск в вебе) в области методов представления и обнаружения знаний, технологий создания и поддержки электронных публикаций, моделирования данных и метаданных.

В последнее время ведется достаточно много изысканий по тематике электронных библиотек. В России, несмотря на обширные теоретические исследования в этой области, результаты которых освещаются, в частности, в научном электронном журнале «Электронные библиотеки» ([www.elbib.ru](http://www.elbib.ru)) и на различных конференциях по тематике ЭБ (например, серия конференций RCDL «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», [www.rcdl.ru](http://www.rcdl.ru)), как правило, законченных практических реализаций («коробочных версий») систем ЭБ мало. За рубежом практических реализаций подобных систем достаточно много, реализуются крупные проекты (например, проекты ассоциации DELOS, [www.delos.info](http://www.delos.info)) по созданию эталонной модели ЭБ и стандартов их разработки. Отчасти это связано с тем, что в разных организациях существуют специфические требования, особенно касающиеся функциональных возможностей отдельно взятой системы. Именно поэтому обычно отдельные электронные библиотеки создаются для специальных приложений и определенных целей и не являются тиражируемым продуктом.

Если говорить о «коробочных решениях» подобных систем, то здесь, в основном, известны такие системы, как DSpace ([www.dspace.org](http://www.dspace.org)), ePrints ([www.eprints.org](http://www.eprints.org)), GreenStone ([www.greenstone.org](http://www.greenstone.org)) и Fedora ([1](http://www.fedora-</a></p></div><div data-bbox=)

commons.org). Также нужно упомянуть об электронной библиотеке «ELSA» (<http://obs.ruslan.ru/?product:ELSA>) – пожалуй, это единственное российское тиражируемое свободно распространяемое решение для электронных библиотек, которое разработано совместно ООО «Открытые библиотечные системы» и Санкт-Петербургским государственным политехническим университетом, хотя аналогичные работы ЭБ также ведутся, например, на базе Ярославского государственного университета им. П.Г. Демидова. Известны и крупные проекты по созданию научных электронных библиотек – научная ЭБ eLibrary ([www.elibrary.ru](http://www.elibrary.ru))<sup>1</sup>, общероссийский математический портал Math-Net.Ru ([www.mathnet.ru](http://www.mathnet.ru)).

Кроме того, недостаточно внимания уделяется особенностям создания ЭБ в вузе – здесь внутри ЭБ естественным образом происходит смешение научных электронных коллекций (ЭК) и образовательных ресурсов, наблюдается гетерогенность источников данных. К тому же могут существовать еще более специфические коллекции документов, например, в Казанском государственном университете (КГУ) за более чем двухсотлетнюю его историю накоплено большое собрание рукописей, старопечатных и редких книг, уникальной периодики. Большая часть этого собрания представляет значительный интерес и используется в учебном процессе и научных изысканиях, однако находится в ветхом состоянии и не выдается читателям на руки. Поэтому наиболее востребованные из этих изданий были переведены в электронную форму, в результате сформировано несколько разрозненных ЭК, основанных на фонде Отдела рукописей и редких книг Научной библиотеки КГУ. При условии, что в организации уже имеются разнородные ЭК, на этапе внедрения одной из существующих коробочных разработок для ЭБ в лучшем случае придется проводить работу по конвертации формата существующих БД ЭК в формат этой системы. В худшем же случае это повлечет за собой создание всех коллекций заново, что приведет фактически к двойной работе и лишним затратам.

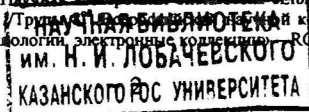
Помимо этого в библиотеках, особенно вузовских, как правило, используются автоматизированные библиотечно-информационные системы (АБИС), которые также являются специализированными хранилищами больших по объему информационных ресурсов.

Таким образом, тема настоящего диссертационного исследования является актуальной.

**Целью работы** является создание принципов построения, моделей и алгоритмов работы научно-образовательной ЭБ.

**Основная задача** – создание модели, алгоритмов и программного обеспечения прототипа системы управления научно-образовательной электронной библиотекой с применением технологий Symantec Web и веб-сервисов,

<sup>1</sup> Глухов В. А., Елизаров А. М. Проект «Научная электронная библиотека eLibrary.ru» и российские электронные журналы: новый этап развития // Труды 4-й всероссийской конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». – RCDL'2006, Суздаль, Россия, 2006. – С. 203-207





объединяющей в единое целое все информационные ресурсы вуза с минимальными изменениями уже существующего массива ЭД. Для достижения поставленной цели было выделено несколько подзадач.

**Основные подзадачи диссертационного исследования:**

- исследование и анализ существующих моделей и реализаций информационных систем управления ЭБ, а также других информационных систем близкой функциональности (электронных архивов, репозиторий и т. п.);
- разработка инфологической модели и логической структуры прототипа системы управления научно-образовательной ЭБ;
- создание моделей выполнения основных операций системы и разработку алгоритмов их реализации;
- реализация компонентов прототипа информационной системы, соответствующей выбранным моделям, методам и технологиям, в виде комплекса программ для ЭБ Казанского государственного университета.

**Методы исследования.** При решении поставленных задач использовались методы математического и инфологического моделирования, системного анализа, информатики, теории информационного поиска и теории баз данных.

**Научная новизна результатов исследований состоит в следующем:**

- построена инфологическая модель системы управления научно-образовательной ЭБ;
- формализованы типовые операции системы ЭБ, разработаны модели и алгоритмы их работы;
- сконструирована логическая структура программного обеспечения (ПО) системы управления ЭБ;
- разработаны компоненты прототипа системы управления научно-образовательной электронной библиотекой.

**Достоверность результатов исследования.** Научные результаты диссертационной работы получены на основании достоверных знаний прикладной информатики, систем управления базами данных, теории информационного поиска и использовании строгого математического аппарата. Полученные результаты подтверждены положительными отзывами о них при обсуждении на международных и всероссийских научных конференциях.

**Практическая значимость** результатов диссертационного исследования заключается в реализации предложенных в работе моделей, методов и алгоритмов при разработке системы управления научно-образовательной электронной библиотекой КГУ. В частности, разработан прототип системы управления вузовской электронной библиотекой, который позволил объединить существующие разрозненные электронные коллекции; автоматизировать процесс создания и предоставления информационных ресурсов пользователям Научной библиотеки Казанского государственного университета по принципу «одного окна».

Внедрение системы управления ЭБ в Научной библиотеке КГУ позволило предоставить единую точку входа ко всем информационным ресурсам ву-

за, что в конечном итоге привело к повышению качества обслуживания читателей библиотеки.

Предложенные инфологические модели, алгоритмы, а также система или ее отдельные модули могут быть использованы в качестве программного решения для электронных библиотек или отдельных электронных коллекций в различных прикладных областях.

**Апробация работы.** Основные результаты диссертации докладывались и обсуждались на объединенном семинаре НИИ математики и механики им. Н.Г. Чеботарева КГУ и факультета вычислительной математики и кибернетики Казанского университета, а также на следующих международных, всероссийских и региональных конференциях: 11-й Международной конференции и выставке «LIBCOM 2007»: Информационные технологии, компьютерные системы и издательская продукция для библиотек (Звенигород, Московская область, 12 – 16 ноября 2007 г.); 3-м Международном форуме (8-й Международной конференции молодых ученых и студентов) «Актуальные проблемы современной науки» (Самара, 20 – 23 ноября 2007 г.); 12-й Международной конференции и выставке «LIBCOM 2008»: Информационные технологии, компьютерные системы и издательская продукция для библиотек (Звенигород, Московская область, 17 – 21 ноября 2008 г.); Международной научной конференции «Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам El-Manuscript-08» (Казань, 26 – 30 августа 2008 г.); 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции – RCDL 2008» (Дубна, 7 – 11 октября 2008 г.); 11-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции – RCDL 2009» (Петрозаводск, 17 – 21 сентября 2009 г.).

**Публикации.** По теме диссертации опубликовано девять научных работ, в том числе одна статья в журнале из списка, рекомендованного ВАК РФ («Ученые записки Казанского университета»).

**Структура и объем работы.** Диссертационная работа состоит из введения, трех глав, заключения, списка литературы, приложения и изложена на 154 страницах.

Исследования проводились в 2007 – 2009 гг. при частичной финансовой поддержке Российского гуманитарного научного фонда (проект 07-01-12146).

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы исследования, описаны решаемые проблемы, рассмотрены общие подходы к построению электронных библиотек, электронных коллекций и интеграции информационных систем. Введение дает характеристику основных проблем и задач, возникающих при этом.

**Первая глава** посвящена анализу основных понятий, задач и функциональных возможностей ЭБ как класса информационных систем.

В настоящее время стандартизованного определения электронной библиотеки не существует. Наиболее адекватным нам представляется определение Т.В. Ершовой и Ю.Е. Хохлова<sup>2</sup>, которые определяют ЭБ как распределенную информационную систему, позволяющую надежно сохранять и эффективно использовать разнородные коллекции электронных документов, а также предоставлять к ним доступ в удобном для пользователя виде через глобальные сети передачи данных.

В процессе становления электронных библиотек были выделены (см., например, итоговый отчет по европейскому проекту DELOS<sup>3</sup>) три типа систем, которые объединяют все необходимые компоненты электронной библиотеки в единую структуру: это собственно электронная библиотека (Digital Library, DL), Информационная система электронной библиотеки (Digital Library System, DLS) и Система управления ЭБ (Digital Library Management System, DLMS).

**Электронная библиотека (DL)** – это система, которая собирает и хранит в течение длительного времени различные наборы электронных документов, в соответствии с принятыми политиками безопасности и правами доступа предоставляет с определенным измеряемым качеством сообществам пользователей специализированный функционал, связанный с этим содержанием.

**Информационная система электронной библиотеки (DLS)** – это система программного обеспечения, которая основана на определенной, в т. ч. и распределенной архитектуре и обеспечивает все функциональные возможности, требуемые ЭБ. Пользователи взаимодействуют с ЭБ через соответствующую систему электронной библиотеки.

**Система управления ЭБ (DLMS)** – это совокупность систем программного обеспечения, которая обеспечивает соответствующую инфраструктуру для формирования системы ЭБ. Подобная инфраструктура включает набор основных функциональных возможностей ЭБ, а также функции для управления ими; объединяет дополнительное программное обеспечение, обеспечивающее специализированные или дополнительные функциональные возможности.

На данный момент не существует единых принципов формирования программного обеспечения ЭБ, а соответствие различным требованиям реализуется, как правило, в зависимости от каждого конкретного случая. Подобный подход не удобен, особенно когда речь идет об объединении некоторых систем в единое целое или хотя бы о взаимодействии систем ЭБ между собой. Для преодоления этого недостатка в рамках одного из крупных европейских

<sup>2</sup> Ершова Т. В. Межведомственная программа «Российские электронные библиотеки» [Электронный ресурс] / Т. В. Ершова, Ю. Е. Хохлов // Электронные библиотеки: рос. науч. электронный журн. – 1999. – Т. 2, вып. 2. – Режим доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/ershova>

<sup>3</sup> The Digital Library Manifesto L. Candela, D. Castelli, Y. Ioannidis, G. Koutrika, P. Pagano, S. Ross, H.-J. Schek, H. Schuldt, 2006,

проектов создана типовая эталонная архитектура ЭБ<sup>4</sup>. Рассматриваемая типовая эталонная архитектура представляет ЭБ не как строго выстроенную вертикальную систему, а как набор независимых компонент, каждый из которых выполняет свою функцию. Она представлена на рис. 1; компоненты организованы в три главных уровня, каждый из которых в свою очередь делится на модули.



Рис. 1. Архитектура ЭБ

На прикладном уровне ЭБ сгруппированы компоненты, которые реализуют прикладную логику библиотеки. *Интегратор* объединяет все компоненты, взаимодействующие с различными внешними ресурсами, которые могут быть неоднократно использованы при создании ЭБ. Способность объединять и использовать различные готовые системы и внешние информационные ресурсы – одно из главных требований при построении объединения систем ЭБ. Остальные компоненты, которые реализуют основной функционал ИС ЭБ, также объединены в модули:

- *модуль управления доступом к ИР* – связан с поиском и доступностью новых ресурсов;
- *модуль управления информационным доменом* включает необходимые инструменты, обеспечивающие сохранность ИР;
- *модуль управления библиотекой* реализует инструментарий, необходимый для управления редакторскими и пользовательскими правами доступа и политиками безопасности, а также управление процессами информационного домена;
- *модуль управления пользователями* реализует функционал работы с информацией о пользователях и группе пользователей;
- *модуль представления (интерфейс пользователя)* – это модуль, отве-

<sup>4</sup> DELOS Workpackage 1. D1.4.2 – Reference Model for Digital Library Management Systems, M. Agosti, L. Candela, D. Castelli, N. Ferro, Y. Ioannidis, G. Koutrika, C. Meghini, P. Pagano, S. Ross, H.-J. Schek, H. Schuldt, 2006

чающий за предоставление всех возможностей ЭБ конечному пользователю; модуль не ограничивается лишь интерфейсом пользователя, сюда также входят любые другие интерфейсы, которые требуются для улучшения доступности, удобства и простоты использования содержимого ЭБ.

Далее идет **уровень взаимодействия**, на котором реализуются необходимые механизмы для связи и обмена данными компонентов между собой, иными словами, этот уровень отвечает за взаимодействие слабосвязанных компонентов ЭБ между собой. Его задача – создать единую, стабильную, корректно работающую систему, скрыть от конечного пользователя возможную разобщенность компонентов системы.

Все вышестоящие компоненты зависят от уровня **физической реализации приложения**, который обеспечивает необходимый функционал для их работы. Поскольку различные среды разработки предоставляют различный функционал, то гибкость и расширяемость всех верхних модулей напрямую зависит от выбранной среды реализации приложения.

Если рассматривать электронную библиотеку вуза, то ни одна из названных систем не учитывает особенностей вуза. Сегодня большинство ЭБ – тематические и содержат в основном электронные аналоги печатных изданий, ЭБ же вуза содержит более широкий спектр информационных ресурсов, т. е. является политематической. Это определяется хотя бы тем, что в вузе существует всегда не менее двух направлений деятельности – образовательная и научная, а в классических университетах, объединяющих множество научных направлений, ЭБ однозначно не может быть посвящена единственной тематике. В зависимости от области применения к ЭБ предъявляются различные требования к ресурсам, метаданным, форматам их описания, которые слабо пересекаются между собой, но должны быть связаны в единую ИС.

При условии, что в организации уже имеются разнородные электронные коллекции, на этапе внедрения новой, вертикально выстроенной системы для хранения и управления сложными электронными объектами так либо иначе возникнут проблемы адаптации существующих ЭК и внедряемой системы, которые в конечном итоге приведут к лишним трудозатратам.

Таким образом, существует потребность в разработке модели и создании прототипа вузовской научно-образовательной ЭБ, которая удовлетворила бы максимум потребностей пользователей, стала составной частью комплекса информационных систем вуза, в частности, КГУ, а также позволила объединить в единое целое существующие наработки в этой области.

Во **второй главе** определяются требования к ИС ЭБ вуза, рассматривается инфологическая модель научно-образовательной электронной библиотеки вуза, анализируется архитектура ЭБ, а также определяются и формализуются алгоритмы типовых операций в системе.

Наиболее распространенным способом долговременного хранения информации в настоящее время являются реляционные базы данных (РБД). Как

известно <sup>5</sup>, РБД можно представить следующей моделью:  $M = (\{A\}, \{R\}, \{D\})$ , где  $\{A\}$  – множество атрибутов,  $\{R\}$  – множество отношений,  $\{D\}$  – множество зависимостей.

Модель  $M$  возникает в результате содержательного анализа предметной области. На практике для получения множества атрибутов, отношений и части зависимостей, описывающих предметную область, используется инфологическое моделирование (см., например, <sup>6</sup> и <sup>7</sup>).

Рассмотрим инфологическую модель системы электронной библиотеки (см. рис. 2). Выделим основные сущности. Поскольку ЭБ состоит из электронных коллекций, то разумно выделить сущность «ЭК». Она должна содержать уникальный идентификатор коллекции и ряд атрибутов. Атрибуты коллекции должны отражать особенности описания самой коллекции, общих свойств документов, содержащихся в ЭК, а также связи между документами и коллекцией.

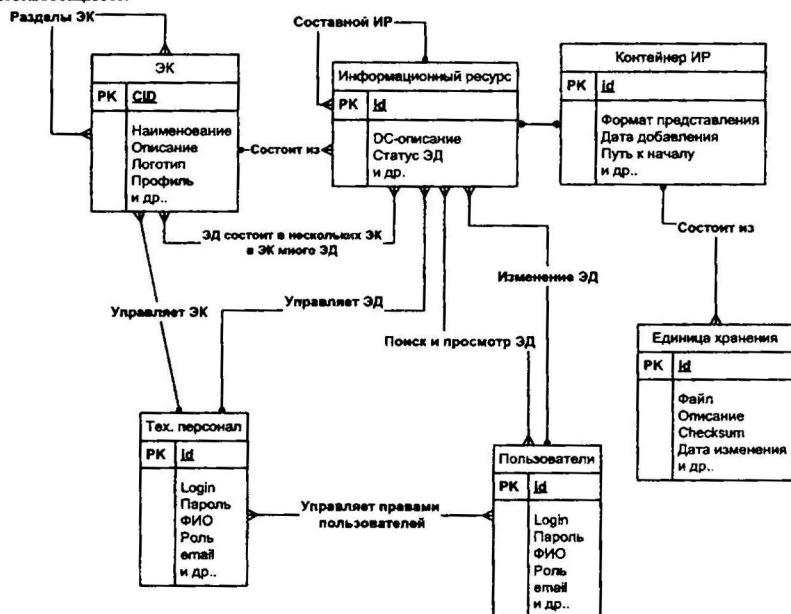


Рис. 2. Инфологическая модель ЭБ

Электронная коллекция состоит из разделов и электронных документов. Раздел коллекции должен содержать в точности такие же атрибуты, что и сама коллекция, поэтому не будем выделять раздел как отдельную сущность,

<sup>5</sup> Кузнецов С. Д. Основы баз данных / С. Д. Кузнецов – М.: Бином, 2007 – 488 с.

<sup>6</sup> Коголовский М. Р. Энциклопедия технологий баз данных / М. Р. Коголовский – М.: Финансы и статистика, 2002 – 800 с.

<sup>7</sup> Чен П. Модель «сущность-связь» – шаг к единому представлению о данных // СУБД – 1995. – № 3. – С. 137 – 158

но должна быть предусмотрена возможность организации древовидной структуры ЭК и их разделов. Информационным ресурсом (ИР) будем называть основную единицу содержания ЭК, ИР состоит из электронного документа и его описания. Соответственно, необходимо рассмотреть сущность «Информационный ресурс». Поскольку в каждой коллекции содержится множество ИР, то здесь присутствуют связи «один – ко – многим». С другой стороны, один ИР может содержаться в нескольких ЭК, т. е. ЭК и ИР связаны связью «многие – ко – многим».

Отдельно охарактеризуем сущность «Информационный ресурс». ИР можно представить как данные (собственно электронный документ) и метаданные, описывающие эти данные. В самом простом случае данные – это бинарный, текстовый или графический файл. С другой стороны, файлов в электронном документе может быть несколько, однако нескольким файлам сопоставляется одно описание ИР. Поскольку ЭБ – это еще и долговременное хранилище данных, то необходимо отслеживать все изменения не только описаний ИР, но и данных документов. Поэтому в рассмотрение введена новая сущность, названная «Контейнер ИР». Каждому описанию ИР ставится в соответствие только один контейнер ресурса (связь «один – к – одному»). Эта сущность помимо уникального идентификатора содержит ряд атрибутов, которые отвечают за целостность и изменение данных документа, т. е. отслеживаются дата добавления/изменения документа, формат представления ЭД, полный путь к данным, связи внутри ЭД. В контейнере ЭД может содержаться несколько «Единиц хранения ИР» («один – ко – многим»). Это сущность, которая содержит информацию о конкретном файле или битовом потоке соответствующего электронного документа (ID, контрольную сумму, связи с другими частями ЭД, описание) и является неделимым информационным объектом. В свою очередь информационный ресурс также может иметь более сложную структуру, например, состоять из различных частей (журнал состоит из статей, книга – из отдельных глав и частей). Здесь возникает вопрос, что считать неделимым ИР. Можно считать один том журнала одним электронным документом. Однако такой информационный ресурс требует внутренней навигации. С другой стороны, каждая журнальная статья имеет своих авторов, поэтому логичнее считать именно отдельную статью неделимым информационным ресурсом. Но статья содержится в журнале и является его частью. Поэтому требуется механизм, который отображал бы подобные иерархии ИР. По этой причине при построении инфологической модели использовано предположение, что ИР может состоять из частей, что позволяет отразить возможную иерархию ИР.

Учитывая текущие тенденции развития СУБД и распространенность реляционных БД (см., например, <sup>8</sup>), нами предложено реализовывать полученную инфологическую модель с помощью реляционной базы данных. Для быстрой и корректной работы инфологическая модель БД должна быть приве-

<sup>8</sup> Кузнецов С. Д. Основы баз данных / С. Д. Кузнецов – М.: Бином, 2007 – 488 с.

дена к нормальной форме. Технические моменты преобразования инфологической модели в нормальную форму и создания реляционной модели БД оставлены вне поля зрения данного исследования, поскольку известны однозначные алгоритмы таких преобразований (см., например, там же).

Условно поделим все ПО для ЭБ на две части (см. рис. 3) – уровень коллекции и уровень электронной библиотеки в целом. На уровне коллекции формируется ПО для отдельно взятой коллекции информационных ресурсов, на уровне библиотек производится объединение всех ЭК в одно целое. Двух-уровневая структура ПО для ЭБ, на наш взгляд, позволяет наряду с созданием новых компонент системы задействовать уже готовые решения, которые могут быть использованы в качестве независимых модулей системы ЭБ. Следовательно, ИС электронной библиотеки имеет сервис-ориентированную архитектуру. Учитывая, что подобная система тесно связана со всемирной паутиной, разработка системы сводится к разработке набора различных веб-сервисов.

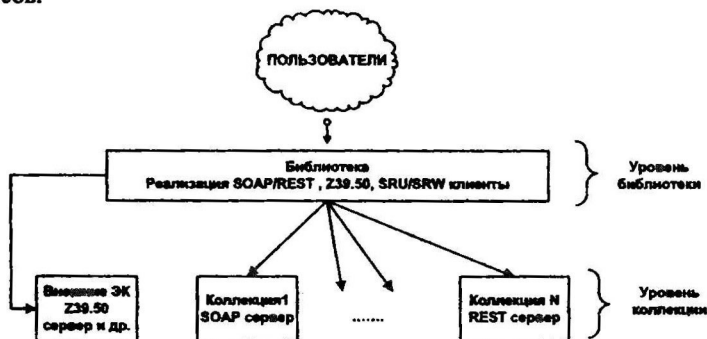


Рис. 3. Логическая структура системы

В диссертации предложены алгоритмы работы с системой. Опишем основные из них.

### Поиск информационных ресурсов

Рассмотрим подробнее операцию поиска документов (рис. 4). Можно выделить простой и общий поиск. Под простым поиском понимается операция поиска данных, выполняемая в рамках одной электронной коллекции. Введем следующие определения.

*Определение 1.* Атомарным будем называть запрос на поиск всех ресурсов, значение определённого атрибута которых удовлетворяет заданному ограничению.

*Определение 2.* Множеством запросов  $Q$  будем называть множество, в которое входят атомарные запросы, выражения вида  $q_1 \wedge q_2$ ,  $q_1 \vee q_2$ ,  $q_1 \oplus q_2$   $q \in Q$ .

Множество запросов мы определяем как множество логических выражений над атомарными запросами. Будем говорить, что ресурс  $R$  удовлетворяет запросу  $q \in Q$ , если справедливо одно из следующих высказываний:

- $q$  является атомарным запросом,  $R$  удовлетворяет  $Q$ ;



- $q$  является выражением вида  $q_1 \wedge q_2$ ,  $R$  удовлетворяет  $q_1$  и  $q_2$
- $q$  является выражением вида  $q_1 \vee q_2$ ,  $R$  удовлетворяет  $q_1$  или  $q_2$
- $q$  является выражением вида  $q_1 \oplus q_2$ ,  $R$  удовлетворяет  $q_1$  и не удовлетворяет  $q_2$

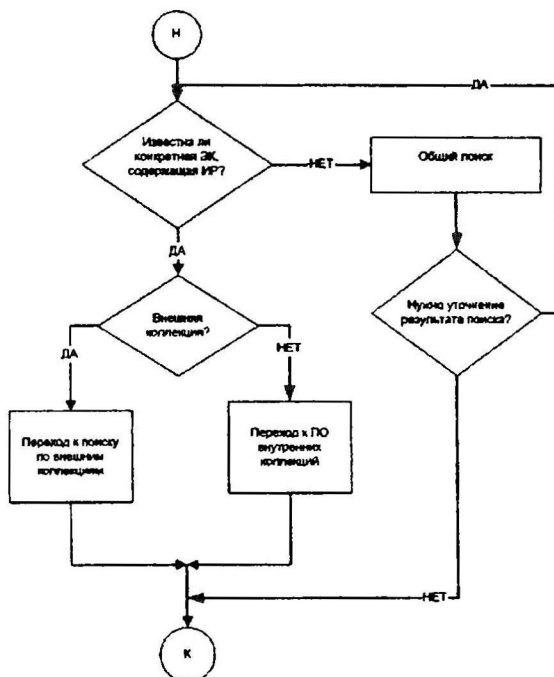


Рис. 4. Схема процесса «Поиск информационных ресурсов»

Тогда операцией простого поиска будет функция  $search(S, q)$ , где  $S$  – схема данных коллекции, которая каждому запросу из множества  $Q$  ставит в соответствие множество удовлетворяющих ему идентификаторов.

**Реализация операции.** Для запросов из множества  $Q$  применяется простой алгоритм построения SQL-запросов для атрибутивного поиска. Для более четкого определения пользователем поисковых термов и минимизации опечаток выполняются запросы для отображения справочных данных из БД в поисковой форме. На следующем этапе из непустых полей формы формируется условие на выборку данных из БД описаний ЭК. Заметим, что должен быть указан хотя бы один поисковый терм, запросы вида «найти все ресурсы» не выполняются.

Под общим поиском будем понимать одновременный поиск по нескольким источникам данных с последующей интеграцией результатов. В общем случае для операции общего поиска задается запрос такого же вида, как и для простого поиска, и указывается коллекция узлов, по которым будет осуществ-

влена операция. В нашем случае в качестве коллекции узлов для поиска выступает все информационное пространство ЭБ. Простой поиск может, с точки зрения пользователя системы, не работать корректно в тех случаях, когда в запросе задаются условия на связанный объект, и удовлетворяющий им связанный объект находится в другом информационном источнике. В таких случаях информации в рамках локального источника недостаточно для выполнения запроса, и часть удовлетворяющих запросу ресурсов будет не найдена.

Если для простого поиска задача выполнения запроса ложится целиком на информационную систему отдельной коллекции, то совместный поиск выполняется при взаимодействии нескольких информационных источников и посредника, роль которого играет программный компонент электронной библиотеки.

*Определение 3.* Источник данных ЭК представляет собой кортеж  $D = \langle i, S, G, m \rangle$ , где  $i$  – идентификатор источника;  $S$  – граф, определяющий схему данных источника;  $G$  – граф, представляющий собой данные источника, соответствующие его схеме  $S$ ;  $m$  – метаданные, соответствующие схеме метаданных коллекции.

*Определение 4.* Под описанием ресурса будем понимать подграф, в который входит сам ресурс, связанные с ним простые и сложные атрибуты, но не входят утверждения относительно связанных с ним классов.

*Определение 5.* Совместимым набором источников  $\Omega$  называется тройка  $\Omega = \langle D, K, w \rangle$ , где:

- $K$  – каноническая схема данных ЭБ;
- $D$  представляет собой множество источников данных  $D = (D_1, D_2, \dots, D_n)$ , таких, что  $\forall D_k, D_l \in D: i_k = i_l \Rightarrow D_k = D_l$ ;
- схема данных каждого источника данных является расширением канонической схемы данных;
- $w$  является функцией, определенной на множестве идентификаторов ресурсов, которая сопоставляет каждому идентификатору из  $G_i$  идентификатор его источника  $i$ .

Для того чтобы можно было отслеживать принадлежность данных, введено понятие соответствия ресурсов и информационных источников. Для каждого ресурса определен один источник, которому он принадлежит, соответственно, данным, полученным из этого источника, для указанного ресурса можно доверять.

*Определение 6.* В объединение данных совместимого набора  $Y$  источников  $\Omega$  входят все тройки, которые являются описанием какого-либо ресурса в источнике данных, если ресурс принадлежит этому источнику.

Тогда под операцией общего поиска будем понимать функцию  $search\_all(\Omega, q)$ , которая каждому запросу из множества  $Q(K)$  ставит в соответствие множество удовлетворяющих ему идентификаторов в  $Y$ .

*Реализация операции*

Общий поиск по электронной библиотеке осуществляется с помощью дублирования поисковых запросов на все серверы, о которых известно сервису электронной библиотеки. Каждая коллекция может быть внутренней или внешней по отношению к сервису ЭБ. В зависимости от коллекции для поиска используются разные веб-сервисы. Поступающие результаты объединяются в одну коллекцию, которая отображается пользователю. Дубликаты в коллекции определяются по совпадению URI.

Поскольку прототип системы реализуется внутри одной организации, то это позволяет пренебречь некоторыми особенностями, которые возникают при работе с распределенной системой.

Таким образом, процесс выполнения запросов можно поделить на несколько шагов.

На первом шаге для каждой коллекции полученный запрос переформулируется так, чтобы программное обеспечение отдельно взятой коллекции могло его обработать. Для этого используется специальное преобразование, изменяющее запрос в зависимости от заданного профиля метаданных коллекции. Например, для передачи запроса по протоколу Z39.50 исходный поисковый запрос должен быть преобразован в type 1 запрос с использованием набора атрибутов BIB-1 протокола Z39.50.

На следующем шаге запрос передается каждой доступной электронной коллекции для выполнения.

На третьем шаге производится сбор полученных ответов. При этом ожидается, что ответы на запросы будут представлены в едином коммуникативном формате, идентификаторы одного и того же документа могут быть отфильтрованы.

Далее производится отображение полученных результатов в виде, понятном пользователю. Учитывая выбор xml-технологий для реализации прототипа ЭБ, пользователю будут представлены результаты в виде xml-файла. Для более удобного отображения к результатам запроса может быть применено xsl-преобразование.

### **Добавление новых ресурсов. Жизненный цикл ИР**

В результате операции добавления обновляются как метаданные, так и данные информационного источника. Ресурс  $R$  можно охарактеризовать его метаданными и данными, т. е.  $R=R(m, d)$ .

*Определение 7.* Операцией добавления ресурса  $R=R(m, d)$  в источник данных  $D=<i, S, G, m>$  будем называть функцию  $add(R, D)$ , которая возвращает новый источник данных  $D'=<i', S', G', m'>$ , такой, что:

$$S'=S;$$

$$M'=m' \cup M, \text{ где } M \text{ и } M' - \text{множества метаданных } D \text{ и } D' \text{ соответственно};$$

$$G'=d' \cup G;$$

$$i \subseteq i';$$

*Реализация операции*

Информационные объекты, прежде чем стать полноценной частью электронной библиотеки, должны пройти несколько стадий опубликования. Опишем жизненный цикл информационных ресурсов (см. рис 5).

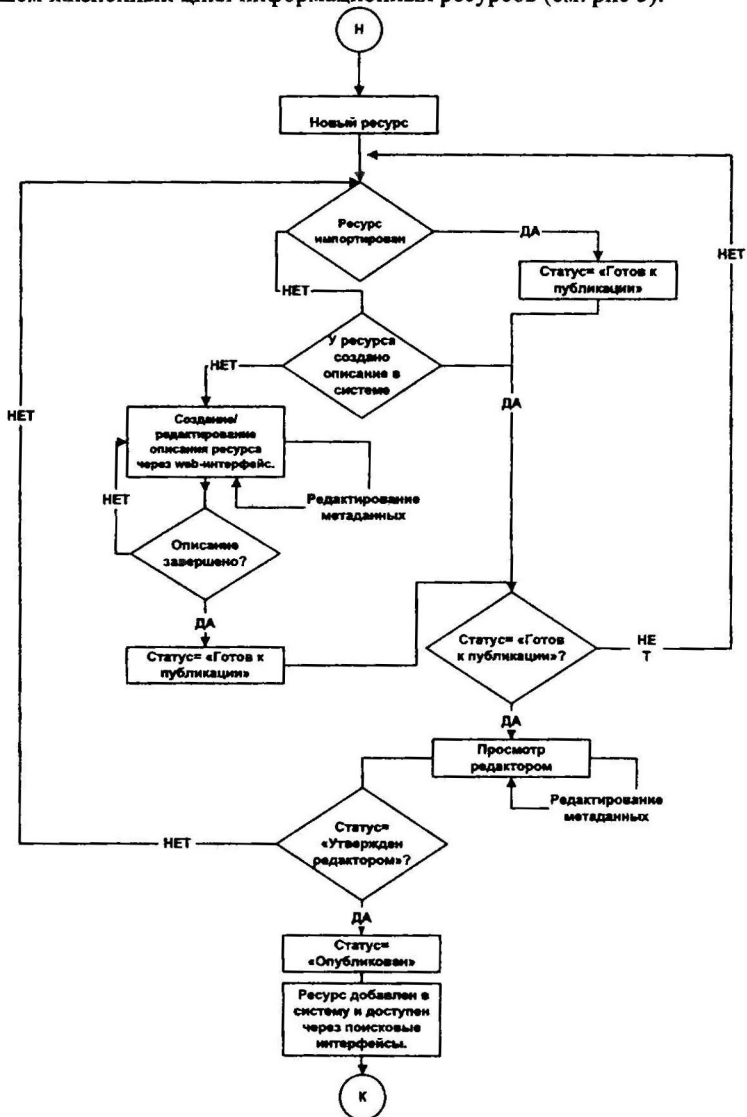


Рис. 5. Схема процесса «Добавление информационного ресурса»

Автор создает новый документ – создается пустой документ, содержащий метаданные нового ИР, он получает статус «Предварительное описание». Далее производится загрузка электронного документа и таким образом создается первичный ИР. Каталогизатор проверяет корректность описания ИР, а также соответствие самого документа и его описания требованиям библиотеки, предъявляемым к электронным ресурсам. Если качество документа не удовлетворительное, то каталогизатор может вернуть документ автору, если же качество ИР удовлетворительное, то ИР присваивается статус «Публичный черновику», при этом может быть проведена необходимая доработка метаданных. После того, как документ получил статус «Публичный черновику», он проверяется редактором и при удовлетворительном результате получает статус «Опубликованный». С этого момента к документу предоставляется публичный доступ, и он становится полноценным элементом электронной коллекции, к нему могут быть применены любые другие операции. В противном случае документ возвращается на предыдущие шаги или вовсе удаляется из коллекции.

Для создания новых документов пользователь должен обладать соответствующими правами доступа. Без проверки редакторами документ хотя и может быть добавлен в систему, но не будет доступен другим пользователям системы. До получения документом статуса «Опубликованный» только автор и редакторы могут производить с ним какие-либо операции.

#### **Создание электронных коллекций**

Поскольку электронная библиотека состоит из коллекций, то необходимо подробнее остановиться на процессах создания новых и подключения внешних коллекций к системе.

Из-за двухуровневой структуры всей системы – уровень коллекции и уровень библиотеки – операция создания коллекций присутствует на обоих уровнях, хотя имеет разные смысл.

*Уровень коллекции.* На этом уровне необходимо сформировать источник данных, иными словами, необходимо определить кортеж  $\langle i, S, G, m \rangle$  и подготовить источник данных к дальнейшему добавлению ресурсов.

В предлагаемом прототипе системы в качестве схемы профиля метаданных используется XML-схема. На основе схемы профиля метаданных определяются остальные множества источника данных.

Таким образом, для создания коллекции необходимо проанализировать схему профиля метаданных. Однако даже все встроенные функции языка позволяют лишь делать частичный анализ профиля метаданных, дальнейшую же обработку и создание структуры базы данных на основе профиля необходимо производить самостоятельно.

Процесс создания коллекции проводится в несколько этапов.

##### *1. Анализ XML-схемы профиля метаданных*

Функции анализа XML-схемы получает на вход файл схемы по указанному пути, производит ее декомпозицию и анализ, в результате чего форми-

руется перечень полей основных таблиц, вспомогательных таблиц (справочников) и их типов.

В процессе разбора XML-схемы используются встроенные средства языков программирования для первичного анализа схемы. С помощью встроенных функций языка файл схемы разбирается и преобразуется в дерево xml-документа, далее производится разбор тегов и отсеиваются лишние значения. Важно учитывать, что нас интересуют лишь теги, имеющие смысловую нагрузку.

### *2. Генерация структуры базы данных на основе профиля метаданных*

Функция создает основные таблицы в базе данных. На вход подается массив с полями и типами полей таблиц, в результате на основе полученных данных генерируется SQL-запрос на создание таблиц.

### *3. Генерация справочников и заполнения их информацией*

На данном этапе справочники создаются и заполняются данными на основе XML-схемы. Заметим, что справочники могут пополняться уже по ходу создания электронных коллекций, однако тогда будет необходим повторный анализ XML-схемы. Программно второй и третий этапы реализованы вместе, но семантически они различны.

На уровне библиотеки смысл операции несколько иной. После создания коллекций необходимо указать сервису библиотеки точки входа для поисковых интерфейсов коллекции. Параметры, вообще говоря, зависят от того, какую коллекцию включаем в библиотеку. Если реализуется SOAP, то необходимо указать wsdl-файл с описанием сервиса, если же речь идет о доступе по протоколу Z39.50, то необходимо указать адрес сервера и список доступных атрибутов BIB-1 и баз данных. Следовательно, создание коллекции на этом уровне – это внесение новой ЭК в список коллекций библиотеки с учетом ее особенностей (профиля метаданных, технологий реализации, протоколов доступа).

### **Регистрация и учет ресурсов**

Электронная библиотека содержит электронные документы, данные и метаданные. Одной из функций подобных систем также является обеспечение сохранности данных системы. Если говорить о метаданных системы, которые хранятся в БД, то их сохранность вполне можно обеспечить инструментами используемой СУБД. Однако кроме метаданных необходимо обеспечивать сохранность и подлинность собственно документов, которые содержит библиотека.

Для построения электронной научно-образовательной библиотеки вуза было принято решение использовать следующий подход – в качестве среды хранения документов использовать файловую систему, в прототипе системы реализовать базовый функционал, который, по крайней мере, мог бы предоставлять информацию о подлинности файла и времени санкционированного и последнего изменения документа и его описания. К тому же, размещение ресурсов в файловой системе веб-сервера также позволяет сделать ресурсы «прозрачными» для поисковых машин интернета.

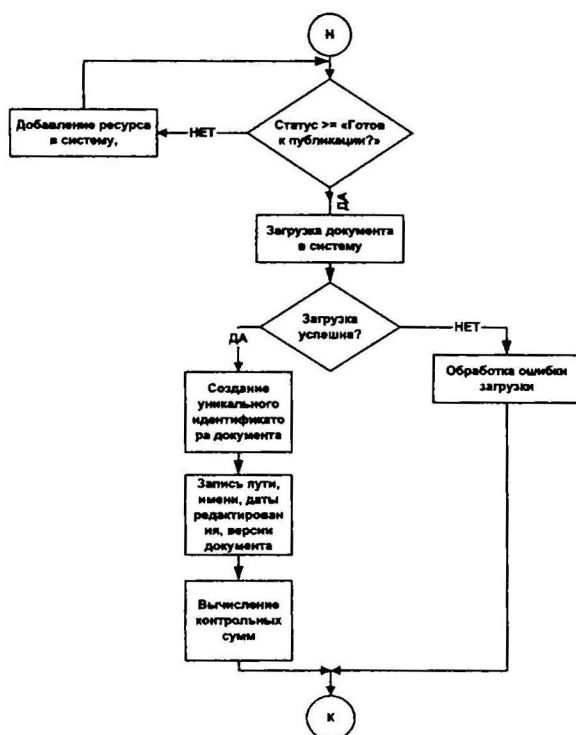


Рис. 6. Схема процесса «Регистрация информационного ресурса»

Схема процесса регистрации ресурсов показана на рис. 6. Опишем его подробнее. После указания статуса «Готов к публикации» (т. е. признака окончания первичного описания документа) система ожидает загрузки документа. В случае успешной загрузки документ будет сохранен на сервере, пути к нему, полное имя и другие данные должны быть зафиксированы в БД коллекции. Документ может принадлежать только одной коллекции, ему в соответствие ставятся его описание и уникальный идентификатор. Для установления подлинности документа вычисляются контрольные суммы, которые также фиксируются в БД. Таким образом, всегда можно установить, изменялся ли документ и когда было последнее изменение.

**Третья глава** посвящена основному применению предлагаемых выше моделей и технологий – разработке компонентов прототипа системы ЭБ.

В настоящий момент в Научной библиотеке КГУ существует два типа коллекций, самым существенным различием между которыми является способ организации: это коллекции, организованные с помощью эксплуатируемой там АБИС «Руслан», и коллекции, ориентированные на использование XML-технологий с профилем метаданных, основанном на Dublin Core. Соответственно, при разработке прототипа системы ЭБ учитывались сложившаяся

ся ситуация и особенности ЭК, предлагаемые решения в первую очередь апробировались на существующих ЭК, делался акцент на максимально полное использование уже существующих модулей, а также готового массива электронных документов. В качестве тестового массива ЭД использовались данные и метаданные коллекции периодической печати, конкретно – номера газеты «Казанские известия» за 1816 год.

На данный момент система представляет следующий функционал:

- *Регистрация, авторизация и изменение профилей пользователей и их прав доступа.* Практически все действия, связанные с изменением ИР или их описаний, не должны быть доступны любому пользователю, поэтому реализованы функции управления пользователями;

- *Добавление ИР, сервисы самокаталогизации, отслеживание жизненного цикла ИР.* Для любого зарегистрированного пользователя доступны сервисы загрузки собственных ИР, каждый ресурс должен иметь корректное описание, поэтому необходимо отслеживать текущие статусы ИР и предоставлять инструментарий для проверки корректности описаний;

- *Регистрация изменений с данными, расположенными на физическом носителе.* Электронный документ случайно или специально может быть искажен или вовсе удален, поэтому реализованы функции, фиксирующие даты изменения документа и проверяющие его подлинность;

- *Создание ЭК на основе формального описания их профиля метаданных.* На основе профиля метаданных ЭК создается поисковая система по данной коллекции, позволяющая производить атрибутный поиск;

- *Поиск и представление ИР пользователям.* Реализованы локальный поиск коллекциям и общий поиск по всей ЭБ, а также отображение результатов пользователю;

- *Передача запроса по протоколу Z39.50 в АБИС и сбор результатов.*

При создании системы управления ЭБ использовалось только свободно распространяемое ПО. Система реализована с использованием XML-технологий на языке программирования PHP, база данных организована на основе СУБД MySQL, также использовались библиотеки Zend Framework.

БД содержит описания ИР, точки подключений и другую вспомогательную информацию, задействованную в процессах обработки и отображения данных. Фактически система состоит из двух слабосвязанных частей, каждая из которых имеет свои БД и программное обеспечение, что отражено в архитектурной модели системы управления вузовской ЭБ. Таким образом, всю программную систему можно разделить на две части в соответствии с уровнями коллекции и библиотеки в целом.

В свою очередь в ПО на каждом из уровней выделено несколько подсистем, реализующих различные группы функций.

*Уровень коллекции.* Программное обеспечение уровня коллекции функционально делится на следующие части:

- подсистема авторизации и управления пользователями;
- подсистема формирования метаданных;



- подсистема поиска в информационном пространстве коллекции.
- На этом уровне библиотеки можно выделить следующие подсистемы:
- авторизации и управления пользователями;
  - управления коллекциями;
  - поиска во всем информационном пространстве библиотеки.

На данный момент поддерживается несколько разновидностей профилей метаданных: Dublin Core с уточняющими квалификаторами и RUSLOM. Для тестовой коллекции периодической печати предложено использовать схему «DC+ уточняющие квалификаторы»<sup>9</sup>. Такой подход является, с нашей точки зрения, более универсальным и позволяет формировать более полные описания, нежели просто элементы Дублинского ядра.

Система предоставляет максимально полный атрибутный поиск по коллекции. При поиске отбираются только те ресурсы, поля которых удовлетворяют заданным атрибутам, незаполненные поля в поиске не участвуют, поля объединяются логической связкой AND.

Поиск возвращает все описания документов, которые удовлетворяют заданным атрибутам и доступны для поиска в виде xml-файла. Для отображения к результатам поиска применяется xsl-преобразование, которое задается при создании ЭК. Таким образом, администратор имеет удобную возможность настройки отображения результатов поиска, что позволяет гибко задавать форматы отображения для каждой коллекции.

Для осуществления общего поиска для коллекций, которые поддерживают технологию веб-сервисов (а таковыми являются все внутренние коллекции), используется парадигма REST. Для связи с коллекциями АБИС реализована поддержка Z39.50 (Z39.50-клиент). При поиске ищутся любые совпадения значений и заданного поискового запроса. Отображение результатов поиска осуществляется в соответствии с квалификаторами Dublin Core в фиксированном табличном виде. Для уточнения полученных результатов предлагается переход к конкретной коллекции, содержащей данное описание.

В заключении приведены основные результаты работы.

В приложении дан краткий обзор некоторых проектов систем электронных библиотек и систем близкой тематики, разработанных в России и за рубежом.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведены исследование и анализ существующих моделей и реализаций информационных систем управления ЭБ, а также других информационных систем близкой функциональности (электронных архивов, репозиторий и т. п.). Проведенный анализ показал их разнородность практически по всем

<sup>9</sup> Абросимов А. Г. Метаданные описания коллекции периодической печати [Электронный ресурс] // Электронные библиотеки: рос. науч. электронный журн. – 2005. – Т. 8 вып. 2. – режим доступа: [http://www.elbib.ru/index.phtml?page=elbib/rus/journal2005/part2/Abrosimov\\_svoobodnyy](http://www.elbib.ru/index.phtml?page=elbib/rus/journal2005/part2/Abrosimov_svoobodnyy)

критериям, по которым он проводился. Обоснована необходимость разработки моделей, технологий и прототипа научно-образовательной ЭБ.

2. Разработана инфологическая модель системы управления научно-образовательной ЭБ вуза, которая учитывает специфику последнего, однако вполне может быть применима для создания любой системы, объединяющей разнородные электронные коллекции.

3. Для научно-образовательной ЭБ предложена двухуровневая логическая структура. Предусматриваемая структура ПО для ЭБ позволяет наряду с созданием новых компонент системы задействовать уже готовые решения, которые могут быть использованы в качестве независимых модулей системы ЭБ, что позволит объединять не только коллекции, созданные самой системой, но и любые электронные коллекции, ПО которых может предоставить соответствующие интерфейсы для связи.

4. Формализованы типовые операции в ЭБ, созданы модели их выполнения и разработаны алгоритмы их реализации, что позволило создать программные модули для реализации данных алгоритмов и системы в целом.

5. На основе рассматриваемых моделей и алгоритмов реализованы программные компоненты и отдельные модули системы управления электронной библиотекой. Разработаны компоненты прототипа системы управления научно-образовательной электронной библиотекой. Предлагаемые подходы, методы, структуры, модели и программные модули практически реализованы в электронной библиотеке Казанского государственного университета.

#### **Основное содержание диссертации изложено**

*В изданиях, рекомендованных ВАК РФ:*

1. Абросимов А. Г., Инфологическая модель научно-образовательной электронной библиотеки вуза / А. Г. Абросимов, Д. С. Зуев // Ученые записки Казанского университета. Сер. Физ.-мат. науки. – 2009 – Т. 151, кн. 3 – С. 167–179.

*В других изданиях:*

2. Зуев Д. С., Программная оболочка для интеграции электронных коллекций Казанского университета [Электронный ресурс] / Д. С. Зуев // Информационные технологии, компьютерные системы и издательская продукция для библиотек «Libcom-2007»: докл. и тез. докл.. – М.: ГПНТБ России, 2007. – URL: <http://www.gpntb.ru/libcom7/disk/12.pdf>.

3. Абросимов А. Г., Принцип построения программного обеспечения электронной коллекции периодической печати / А. Г. Абросимов, Д. С. Зуев // Актуальные проблемы современной науки: труды 3-го Междунар. форума (8-й междунар. конф. молодых ученых и студентов). Естественные науки. Ч. 1–2: Математика. Математическое моделирование. – Самара: Изд-во Сам-ГТУ, 2007. – С. 78–83.

4. Абросимов А. Г., Научно-образовательная электронная библиотека вуза / А. Г. Абросимов, Д. С. Зуев // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: труды 10-й Всерос. науч. конф. «RCDL'2008» (Дубна, Россия, 7-11 октября 2008 г.). – Дубна: ОИЯИ, 2008 – С. 374–379.

5. Абросимов А. Г., Электронная коллекция информационно-образовательных ресурсов КГУ [Электронный ресурс] / А. Г. Абросимов, Д. С. Зуев // Информационные технологии, компьютерные системы и издательская продукция для библиотек: 12-я Междунар. конф. и выставка «Libcom-2008»: докл. и тез. докл. – М.: ГПНТБ России, 2008. – URL: <http://www.gpntb.ru/libcom8/disk/2.pdf>.

6. Зуев Д. С., Принципы построения программного обеспечения для разнородных коллекций в вузе / Д. С. Зуев // Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам: материалы Междунар. науч. конф. (Казань, 26–30 августа 2008 г.) / отв. ред. В. Д. Соловьев, В. А. Баранов. – Казань: Изд-во Казан. гос. ун-та, 2008 – С. 123–127.

7. Зуев Д. С., Модели и принципы построения прототипа программной системы управления вузовской электронной библиотекой / Д. С. Зуев // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: труды 11-й Всерос. науч. конф. «RCDL'2009» (Петрозаводск, Россия, 17–21 сентября 2009 г.). – Петрозаводск: ИПМИ КарНЦ РАН, 2009. – С. 203–209.

8. Абросимов А. Г., Принципы создания электронной библиотеки: опыт Казанского ГУ / А. Г. Абросимов, Д. С. Зуев // Государственная и муниципальная служба в России и Татарстане: истоки и современные тенденции развития: Всероссийская научно-практическая конференция, посвященная 250-летию подготовки кадров для государственной службы в г. Казани: материалы пленарного и секционных заседаний. – Казань: Центр инновационных технологий, 2009. – Кн. 3. – С. 18-24.

9. Зуев Д. С., Прототип научно-образовательной библиотеки вуза / Д. С. Зуев // Труды Математического центра имени Н. И. Лобачевского: Материалы Восьмой молодежной научной школы-конференции «Лобачевские чтения – 2009» (Казань 1 – 6 ноября 2009 г.) – Казань: Казан. матем. об-во., 2009. – Т. 39. – С. 232-235.

В заключение считаю необходимым выразить благодарность своему научному руководителю профессору А. М. Елизарову, А. Г. Абросимову, а также сотрудникам Научной библиотеки им. Н. И. Лобачевского Казанского государственного университета за неоценимую помощь, оказанную при написании диссертационной работы.



Отпечатано с готового оригинала-макета  
в типографии издательства  
Казанского государственного университета  
Тираж 100 экз. Заказ 95/2

420008, ул. Профессора Нужина, 1/37  
тел.: (843)233-73-59, 292-65-60







